



## 検索戦略と最適化

2015  
時実 象一  
tokizane@aichi-u.ac.jp



### 検索の進め方 (1)



- ツールを選ぶ
  - 検索目的にあったデータベース
    - 主題対象
      - 新聞記事、判例、学術文献
    - 収録範囲
      - 国、年代
    - 収録情報
      - 書誌、抄録、索引、本文、データ

3

### 検索の進め方 (2)



- 概念をまとめる
  - 最大3つくらい
  - 例:老人のスポーツ推進活動
    - 「老人」
    - 「スポーツ」
    - 「推進活動」

4

### 検索の進め方 (3)



- キーワード (同義語) を選ぶ
  - 「老人」→老人、高齢者、年寄り、シニア
  - 「スポーツ」→スポーツ、運動、テニス、ゴルフ、ゲートボール、水泳、マラソン
  - 「推進活動」→推進、振興、運動、活動、施策、政策、自治体、市、町、ボランティア
- 複合語はなるべく使わない
  - 「老人のスポーツ」→「老人」AND「スポーツ」

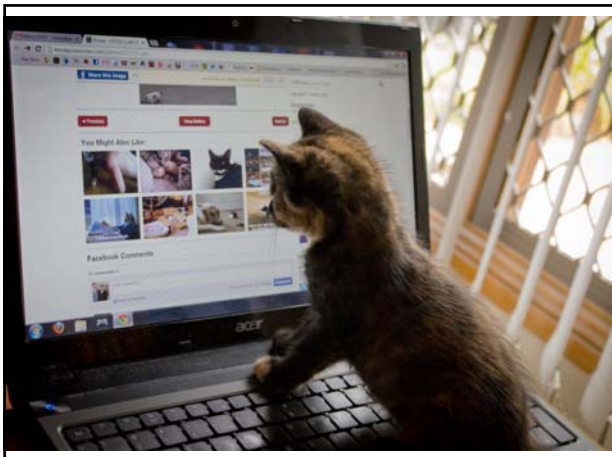
5

### 検索の進め方 (4)



- 段階的に検索して評価
  - たとえばまず「老人」の概念と「スポーツ」の概念で検索してみる
  - 結果がよければそこで終了
  - 結果が悪ければ改良 (最適化)

6



### 検索式の改良 (最適化)



- 検索を広げる - 漏れを減らす
  - 同義語を追加 (OR) (老人 → 高齢者)
  - より広い語に置き換える (キャベツ → 野菜)
  - より広いフィールドを使う (論文名 → フリーワード)
- 検索を絞る - ノイズを減らす
  - 絞る語を追加 (AND)
  - より狭い語に置き換える
  - フィールドを限定
  - 除外する語を指定 (NOT) (気をつけて使う)

8

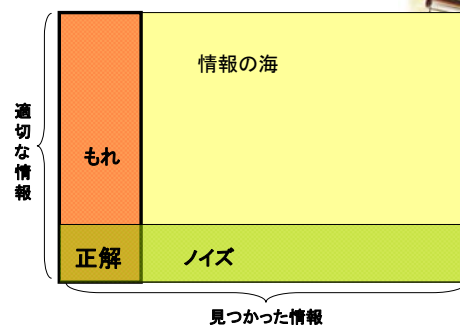
### 検索結果の最適化



- 再現率と適合率 (精度)
- ランキング

9

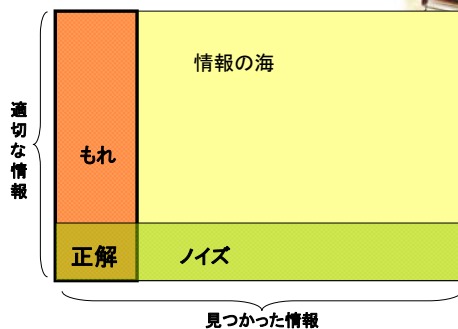
### 「もれ」と「ノイズ」



• もれの大きさは正確にはわからない

10

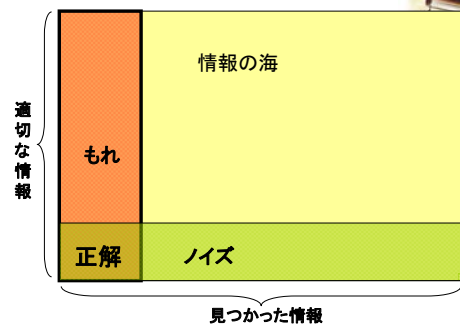
### 再現率 (正解/正解+もれ) 見つかった情報の割合



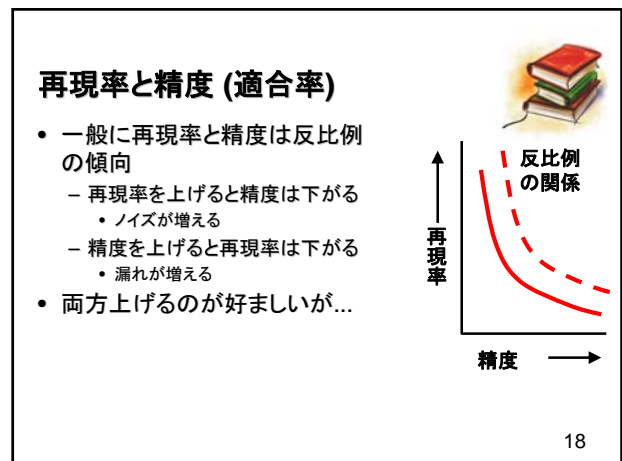
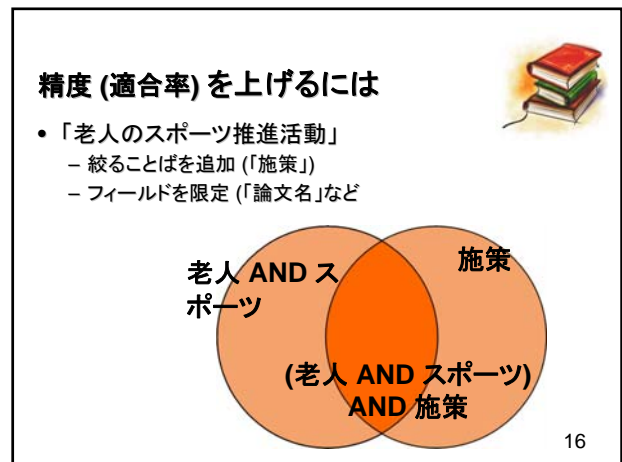
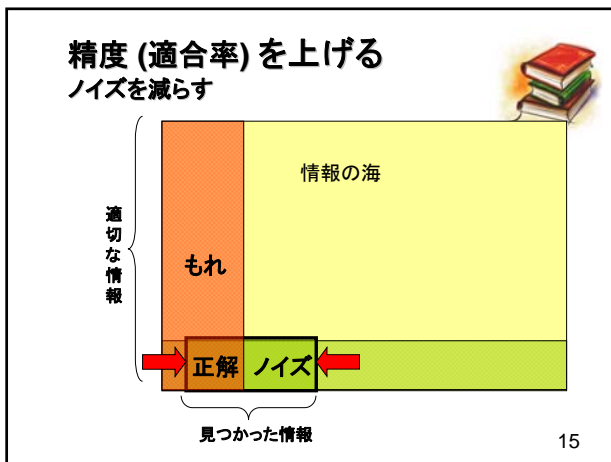
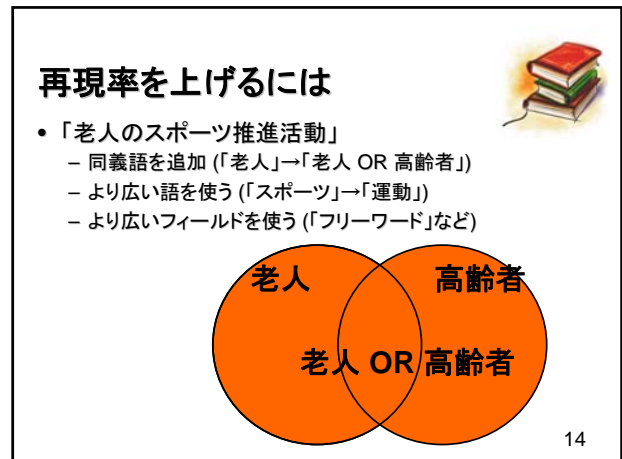
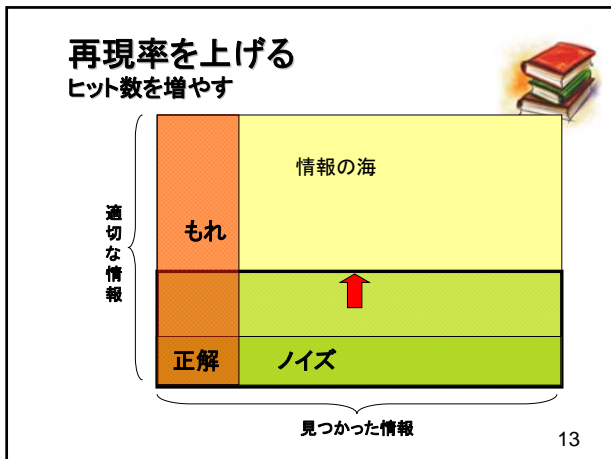
• もれの大きさは正確にはわからない

11

### 精度 (適合率) (正解/正解+ノイズ) 欲しい情報の割合



12



### 再現率と精度 (適合率)



- 「老人のスポーツ推進活動」(CiNii, 2014)
  - タイトルで検索、リハビリ関係は除く

A	B	C	ヒット	正解	ノイズ	再現率*	適合率 (精度)
老人	スポーツ						
老人	スポーツ	推進					
老人 OR 高齢者	スポーツ	推進					
老人 OR 高齢者	スポーツ	推進 OR 政策 OR 施策 OR 行政 OR サークル OR イベント					

19

### 検索結果の配列 (sort)



- 時系列順
  - 通常新しい順 (新聞記事、CiNii, NDL-OPAC)
- 五十音順 (OPAC など)
- 適合順 (Ranking)(検索エンジン)
  - 検索結果をもっともな (適合しているとおもわれる) 順に配列する

20

### 検索範囲の制限



- 検索をある範囲に制限
  - 年代
  - 資料種類
  - 言語
  - ドメイン (検索エンジン)
  - ファイルタイプ

21

### ランキング



- 適合する情報とは
  - 検索キーワード
    - 一致
    - 近接
    - キーワードのある場所
  - 利用者が欲しい情報

22

### ランキング



- 欲しい情報とは
  - 重要な情報
    - 公的なサイト
    - 他からのリンクが多いサイト
  - 人気の高い情報
    - アクセス数が多いサイト
    - 他からのリンクが多いサイト
  - 信頼性の高い情報
    - 公的なサイト
    - 他からのリンクが多いサイト

23

### ランキング



- PageRank (Google)
  - リンクが多いサイトは点数が高い
  - 点数の高いサイトからのリンクは重みが高くなる

24

### ランキング



- 検索エンジン最適化
  - Search Engine Optimization
  - 自分のホームページが検索結果の上位に来るように最適化する
  - 商売では重要

25

